Computing value per square foot of vacant parcels and aggregating to model zone level *

Huiling Zhang †and Richard Arnott

Department of Economics,
University of California, Riverside

Abstract

Average value per square foot of vacant parcels within a model zone has been of interest. Given the fact that a parcel's assessed value is a fixed fraction of its last-sale price, all values should be expressed in 2000 year dollars in order to be comparable. OLS regression of natural logarithm of value per square foot of vacant parcels on accessibility variables, city dummy, planned land use dummy, and last-sale year dummy permits imputing missing values.

Introduction

A model zone's average value per square foot of its vacant parcels equals to total value of its vacant parcels divided by total area of its vacant parcels. Area of any parcel can be easily found with the help of a GIS software, but parcel value, which was collected and processed separately by assessor offices of each county, contains lots of missing data. If the parcels with missing values are drawn from the same distribution in terms of value per unit of land area as the parcels with values given, or less restrictively, then simply averaging over parcels with values given provides an unbiased estimate of the value per square foot of vacant parcels across a model zone. Actually, the estimate is unbiased as long as the mean value per unit of land area for the two sets of parcels are equal. However, it is not supported by data. **How can we test this?**

A more plausible assumption is that the relationship between land value per square foot and characteristics is the same for the two sets of parcels. **How can we justify this?** Then missing land values can be estimated from observable characteristics. Averaging over given land values and estimated land values generates an unbiased estimate even if missing land

*This report was prepared in conjunction with the University of California MRPI Project LA-Plan.

†Please address questions and comments to the primary author, Huiling Zhang, at hzhan017@ucr.edu.

values exhibit a different distribution from given land values. The relationship reveals how underlying geographical and physical characteristics affect land values. And the value of land lies in its capacity or ability to serve human interests. Land use type was classified by, as the name itself suggests, the way a piece of land or the structure built on it serves human interests. Therefore it is reasonable to assume that the same relationship holds within a land use type, or even across similar land uses.

This sub-project deals with vacant land. The concept of vacant land here is a broad one, it includes five major categories (see table 2): under construction, open space and recreation, agricultural, vacant without any structure built on it, and former military base. **justify why those land uses can be classified as vacant** The detailed classification of parcels was provided by Southern California Association of Governments (SCAG). It gives every parcel a current land use code and planned land use code based on regular assessments. And this sub-project used the land use classification in 2008. [1]

It is important to make the distinction between developable vacant land and undevelopable vacant land. This sub-project aims at providing initial numbers of property values and rents for the RELU-TRAN model in MRPI project. The RELU-TRAN model is a general equilibrium one, in which prices of different factors interacts and affects each other. The value of a piece of vacant land that can be developed into other land uses should be affected by the market value of other land uses. While a vacant land that provides only aesthetic values should not have much correlations to the market values of other land uses, or be affected in a different way.

This sub-project estimated the relationship between natural logarithm of assessed value per unit of land area of vacant land against geographical and physical characteristics of the land. Then missing land values were estimated. Finally value per square foot of vacant land for each model zone is calculated as the average of given land values and estimated land values over total land area.

This report has five sections. **changing** Section I documents data source and data preparation methods. Section II provides regression results and imputation methods, including how to discount value per square foot into year 2000. Section III documents how to aggregate parcel information into model zone level. Section IV documents how to present results on a map, and Section V concludes.

1 conceptual framework

There is an extensive literature on hedonic pricing of land and residential properties. The classic model says that a property's value is determined by its access to amenities, services and jobs as well as its physical characteristics.

From available data, we can estimate the relationship between land value and various contributing factors. If the relationship also applies to land of which land value data is unavailable but accessibility variables and other contributing factor variables are available,

[1] Why do we use lu08, b/c it is the most recent one?

There is no data on the physical characteristics of vacant lands, except for land sizes. But there is data on the type of vacant land. Therefore a dummy variable on current land use of a piece of vacant land can be added into the regression equation to capture the average physical differences across types.

$$\ln(vsq) = \beta_1 X_1 + \beta_2 X_2 + \gamma_1 saleyr + \gamma_2 city + \epsilon \quad (1)$$

A regression reveals how value per square foot of vacant parcels was determined by various factors. In this project, accessibility measures, including distances to the nearest freeway, coast, sub-centers and CBD, planned land use, year of last sale, and the belonging city of a vacant parcel acted as explanatory variables.

Planned land use is special to vacant parcels compared to other types of parcels. A cropland that is going to be transformed into residential houses will be more valuable than if it is to remain as cropland. And transforming into residential housing would be more valuable than transforming into public parking lot. So planned land use as a categorical explanatory variable should be in the regression equation.

Values should all be discounted to year 2000 for them to be comparable.[2] Since the value used was "TOTVAL07", whose value was calculated by county assessor offices by multiplying a constant fraction on the last sale price. So it was largely affected by the year when its last sale occurred. A piece of vacant land that has earlier last sale date would be more severely underestimated in value. So inflating or deflating the assessed values to a reference year, in this project 2000, is a must. This is why last sale year should enter the regression equation as a dummy explanatory variable.

Vacant land's value has also been affected by the city it was located in. Cities vary in their infrastructure level, economic policies, public services, environment quality and etc. Two pieces of vacant land with all qualities the same except for the city they were located in may possibly differ in value per square foot. So city should be one of the categorical dummy variables too.

2 Data and Variables

Five primary data sources contributed as data input to this sub-project. The first three are about parcels, and the last two are useful in calculating accessibility measures. The following table documents the variables of use in this project from different data sources.

[2]TOTVAL07 is the assessed value in 2007. Due to Proposition 13 (California's property tax revolt Proposition) assessed value is not the assessor's estimate of true market value in 2007, but instead the property tax base of the parcel in 2007. The property tax base of vacant parcels in 2007 is calculated as the most recent sales price compounded annually by the maximum of 2% and an administratively-decided inflation from the date of the most recent sale to 2007 (the procedure is more complicated for developed properties since account is taken of improvements). Since TOTVAL07 is therefore a function only of the most recent sales price and the year of the most recent sale, a property's market value in 2000 can be estimated by using TOTVAL07 and a time dummy for the year of most recent sale.

Table 1: Input data files and their locations in ftp server

Data Contents	Useful Variables	Source
Parcels with model zone code shape file	"SCAGXYID, Shape_Area, X-Y coordinates, MZ and APN"	(ftp address `ftp://mrpi.geog.ucsb.edu/data2010/parcel/MZparcel_imputed_GIS/`)
Assessed value of parcels table	"SCAGXYID, APN, TOTVAL07, City_name, and SCAG_GP_CO"	(ftp address`ftp://mrpi.geog.ucsb.edu/data2010/parcel/attribute_whole/`)
Latest transaction records of parcels	"APN, and DT_SALE"	(ftp address`ftp://mrpi.geog.ucsb.edu/data2010/parcel/raw/`)
Freeway and coastline shape files		
Employment sub-centers and CBD coordinates		

The correspondence between variable name and attribute field name are also in the table. [3]

As for the table [4]

In the regression, dependent variable is natural logarithm of value per square foot, which equals assessed value ($TOTVAL07$) scaled by area ($Shape_Area$). Independent variables are accessibility variables, which are calculated from the X-Y coordinates of parcels, sub-centers, and CBD; and shape files of freeway and coastline. $SCAGXYID$ and APN were keywords in matching parcels from different data sources. Model zone codes (MZ) were used in identifying parcels that are from the same model zone.

3 Data preparation

All data preparation work was done in ArcCatalog.

3.1 Developability...

Developability of a vacant parcel is decided according to a procedure generated by Guo and Arnott. Table 2 displays the developability of every land use.

[3] To ease calculation burden, it is recommended to export only useful attributes of any file, no matter it is shape file or dbf table or csv table.

[4] Add a column called "Corresponding Attribute fields in DBF file"

3.2 Select vacant parcels out

Vacant parcels in this sub project refers to parcels that have land use code[5], either LU08 or LU_08, that falls into one of the three major categories: 1700, 1800, and 2000. Data inputs of this project share the characteristic that data size is huge. Because only vacant parcels are of interest in this sub project, selecting vacant parcels out was done before any process was taken to avoid unnecessary calculations.

In data source 1, parcel model zone shape file, land use code is LU_08, while for data source 2, assessed value table file, land use code is LU08. [6]

LU08 and LU_08 were prepared in different methods, it was consensus of the project team that LU08 should be relied upon when deciding the land use type of a certain parcel. But random sample comparison revealed that the two codes were not significantly different in identifying the same parcel, so for the sake of easing calculation burden, both codes were used.

3.3 Calculate accessibility measures

Four kinds of accessibility measures used in this sub-project can be grouped into two according to their calculation methods.

- Distance to the nearest freeway and to the nearest coast:
  They are point to line distances, and can be done only in shape files. "near" function in proximity of analysis tools of ArcMap was employed.

- Distance to the nearest sub-center and to CBD:
  They are point to point distances, so the calculations are done in R codes, detail of which can be found in Appendix I.

3.4 Combine parcels' information from different data sources

"Join" function of ArcCatalog was employed. Join 2 to 1 using SCAGXYID, and join 3 to 1 using APN.Output: LA_join2.dbf. (scagxyid, $shape_area$, $x - ycoordinates$, apn, fsub, cbd, freeway, ocean, totval07, $city_name$, scag_gp_co, and dt_sale.)

3.5 Merge six counties' data

"Merge" function of ArcCatalog was employed. Notice one small thing about extracting "last_sale_year" from dt_sale. Different counties have different format recording last-sale date.

---

[5]For detailed interpretations of land use classification and code system employed in this sub-project, please refer to `ftp://mrpi.geog.ucsb.edu/data/11_Land_use/LU_CODE/`

[6]Query syntax in ArcGIS is:
"LU_08" >= 1700 AND "LU_08" < 3000 AND "LU_08" <> 1900
"LU08" >= 1700 AND "LU08" < 3000 AND "LU08" <> 1900
Notice that both codes system assign same number to same type of parcel.

Table 2: Land uses that were treated as vacant and their corresponding developability

Main categories	Special type	Developable
1700 Under Construction		Y
1800 Open Space and Recreation	1810 Golf Courses	Y
	1820 Local Parks and Recreation (1990 Database only)	N
	1821 Developed Local Parks and Recreation	N
	1822 Undeveloped Local Parks and Recreation	N
	1830 Regional Parks and Recreation (1990 Database only)	N
	1831 Developed Regional Parks and Recreation	N
	1832 Undeveloped Regional Parks and Recreation	N
	1840 Cemeteries	Y
	1850 Wildlife Preserves and Sanctuaries	N
	1860 Specimen Gardens and Arboreta	N
	1870 Beach Park	N
	1880 Other Open Space and Recreation	N
2000 Agriculture	2100 Cropland and Improved Pasture Land	Y
	2110 Irrigated Cropland and Improved Pasture Land	Y
	2120 Non-Irrigated Cropland and Improved Pasture Land	Y
	2200 Orchards and Vineyards	Y
	2300 Nurseri	Y
	2400 Dairy, Intensive Livestock, and Associated Facilities	Y
	2500 Poultry Operations	Y
	2600 Other Agriculture	Y
	2700 Horse Ranches	Y
3000 Vacant	3100 Vacant Undifferentiated	Y/N
	3200 Abandoned Orchards and Vineyards	Y
	3300 Vacant With Limited Improvements	Y
	3400 Beaches (Vacant)	Y
1270 Military (Vacant)	1274 Former Base (Built-up area)	Y
	1275 Former Base (Vacant Area)	Y
	1276 Former Base Air Field	Y/N

Note: Developability of a land use type was assigned according to a procedure generated by Guo and Arnott. For land uses with code as 3100 or 3000, a spatial smoothing technique was employed to determine developability. For other land uses, developability was fully determined by the land uses.

Table 3: Data and Variables

	Data Source	Useful Variables
1	"LA_MZ_dist.dbf"	scagxyid, shape_area, x-y coordinates, apn, fsub, cbd, freeway, ocean
2	"LAX_ATT.dbf"	scagxyid, apn, totval07, city_name, and scag_gp_co
3	"last_sale_LA.dbf"	apn, dt_sale

3.6 Blanks and zeros in datasets

In this sub-project, datasets used are flawed in many different ways. So blanks and zeros were all treated as missing values. But this should not significantly alter regression and imputation results on the aggregated model zone level.

4 Regression and Imputation

Response variable:

$\ln(vsq)$: Natural logarithm of value per square foot of a vacant parcel

Explanatory variables:

X_1: Accessibility variables: fsub, CBD, freeway, and ocean

X_2: planned land use dummies: *scag_gp_co*, default single family residential (1100) [7]

last-sale year dummies: default 2000

city dummies: default Los Angeles

Regression sample:

Parcels that have data on lgvsqft, accessibility variables, and last-sale year. Notice that Imperial County has no data on last-sale year, therefore not a parcel of it was included in the regression.

5 Report regression results...

Discuss the statistical significance, sign, and magnitude of coefficients on accessibility measures.

Discuss the joint statistical significance of city dummies.

Discuss the joint statistical significance of last sale year dummies.

Discuss model specification. For example, whether the error term exhibits normal distribution.

[7] For documentation on correspondence between planned land use and planned land use code assigned, please refer to `ftp://mrpi.geog.ucsb.edu/data/10_Parcel/GP_LU_Correspondent.doc`

5.1 Discount to year 2000

For parcels that have data on both last-sale year and assessed value totval07, subtracting regression coefficients of its last-sale year from lgvsqft yields the natural logarithm of value in 2000 dollars.

5.2 Imputation

Impute value per square foot in 2000 dollars for parcels with missing last-sale year or assessed value totval07. Set all last-sale year dummies to zero and apply regression coefficients produces predicted value.

This part was done in STATA, and codes can be found in Appendix II.Output: lgvsqft_adj

6 Aggregate to model zone level

Average value per square foot in 2000 dollars of vacant parcels of a model zone is calculated in this way:

First, get the value in 2000 dollars of every vacant parcel of a model zone. If the land value and last-sale year of a vacant parcel is available in the database, then the value here is raw or true data. Otherwise, imputed values were applied.
Second, add up adjusted values of all vacant parcels belonging to one model zone.
Third, add up area of all vacant parcels belonging to one model zone.
Finally, total adjusted values scaled by total area would yield adjusted value per square foot of vacant parcels in a model zone.

This part was done in ArcMap using function "summarize" by model zone in field calculator of attribute table.

7 Mapping

To provide an intuitive presentation of average value per square foot of vacant parcels within a model zone, colored map of Los Angeles Metropolitan Area was used. The darker a model zone is on the map, the higher value per square foot of its vacant parcels has on average.[8]

Blank areas in the map represent parcels that were lost in the process of producing model-zone-based map from parcel-based map, which was done by tool "dissolve" in ArcGIS software. Fortunately, only some model zone in Los Angeles County suffered this problem, and there is no model zone that lost all of its parcels.

This problem may take a while to be fixed, due to some unknown technical reason. As the data loss occurred only at the phase of producing maps, it did not affect regression and imputation results at all. Therefore, it was left for future treatment.

Four steps were taken before creating the map.
Export: parcel model zone shape files are too large in size, so select fields from attribute

[8]This part was done in ArcMap in "Symbology" Panel of "Layer Properties".

table is a good way to speed up calculation time. In this case, scagxyid and model zone code are the only two fields that are necessary.
Dissolve, Merge, and then a final Dissolve.

8 Developable and Undevelopable

A piece of vacant land is undevelopable if it can not be developed according to federal and state laws or the development cost is prohibitively high due to physical characteristics. For example, national parks are not developable. Developability of a vacant parcel is assigned due to a procedure designed by Ross Guo and Richard Arnott, and the associated technical report can be found in `http://vcpa.ucr.edu/Papers.html`.

The objective of this sub-project is to generate value per unit of floor area at model zone level to initiate running the RELU-TRAN model. And only developable vacant land of is of interest in the model. Therefore, developable vacant land and undevelopable vacant land are treated separately in this sub-project. Value per unit of floor area for undevelopable land is also calculated to serve as a robustness check of the imputation procedure applied to developable vacant land.

Two separate regressions were run for developable and undevelopable land, because there is no reason to assume that accessibility affects undevelopable and developable in the same way. Table 4 compares the average per unit area value of vacant parcels by developability. Generally developable land are more valuable than undevelopable land, but some exceptions exist, like model zone 52 "South Buena Park". The reason might be that developable land were mis-classified as undevelopable, either due to imperfections in classification procedure design or coding error of land use type in SCAG parcel database.

9 Discussion

Places where improvement can be achieved given more time:

9.1 Detect and adjust spatial autocorrelation of parcel data

Spatial autocorrelation should be taken care of if regression data is spatial. If an autocorrelation pattern is detected, all the regression results and imputed aggregated average value per square foot for model zone may be false. But it would just be a change of parameters and a simple re-run of ready-made codes.

Several empirical paper [1] have found that the regression results are not much different with or without spatial autocorrelation adjusted.[9]

Although spatial autocorrelation should not be neglected, it may have limited adverse effects on the reliability of estimated values at model zone level. In this project, regression equations include city dummies. If spatial autocorrelation of land values only exists within

[9] Did they give potential explanations for this?

a city boundary, then the estimated coefficients with city dummies should be reliable even without adjusting spatial autocorrelation. But it is too restrictive to assume that the value of bordering lands are not related. However, a model zone usually covers several cities, the correlation of land values across model zones should be much smaller given that spatial autocorrelation is more of a local concept.[10]

9.2 Make use of last-sale price data

Price of last-sale is the ideal variable for measuring value of a parcel, but as a result of the low quality of its data, assessed value "totval07" was used instead. Orange County and Riverside County have data on both variables, which provides an opportunity to compare the results using two different dependent variables.

9.3 former military bases and parcels with land use code as "0"

9.4

Acknowledgements

The authors would like to thank the following persons for their constant support:
Wenwen Li, for her prompt and detailed answers to whatever technical questions and data questions; Yizhen Gu for his specialized knowledge in the data; and Jifei Ban for his well-developed procedures and codes for solving the problems.

References

[1] Conway, D., Li, C., Wolch, J.et al. (2008) A spatial autocorrelation approach for examining the effects of urban greenspace on residential property, Journal of Real Estate Finance and Economics, $10.1007/s11146-008-9159-6$.

Appendix

Table 4: Estimated value per unit floor area of **vacant land** by developability

Model Zone		Developable		Undevelopable	
ID	Name	# of parcels	Value per unit lot size($\$/ft^2$)	# of parcels	Value per unit lot size($\$/ft^2$)
1	Downtown Los Angeles	725	108.56	119	107.02

[10] provide literature support here, though Prof. Goodchild said so in one of the meetings.

Model Zone		Developable		Undevelopable	
ID	Name	# of parcels	Value per unit lot size($\$/ft^2$)	# of parcels	Value per unit lot size($\$/ft^2$)
2	Westside	51	70.44	17	78.08
3	Glendale	2696	37.15	167	30.03
4	East Los Angeles	2456	30.60	57	34.41
5	Maywood	15	20.89	43	61.88
6	Florence	41	66.52	53	46.52
7	Baldwin Hills	379	148.83	112	101.33
8	Beverly Hills	1764	124.54	211	114.99
9	El Segundo	544	112.02	536	279.34
10	Santa Monica	87	101.52	74	156.06
11	Marina del Rey	179	45.47	53	96.44
12	Westwod	8	85.58	2	171.55
13	East Santa Monica Mtns	3795	108.81	1397	53.69
14	Reseda - van Nuys	585	47.85	45	45.83
15	East van Nuys	170	57.50	45	53.79
16	Burbank	441	26.14	1850	8.71
17	Pasadena	1255	31.76	90	31.18
18	East Pasadena	178	34.25	233	36.66
19	Rosemead	288	22.60	110	24.23
20	Pico Rivera	211	57.29	84	51.22
21	South Gate	229	55.97	88	52.13
22	West Compton	243	50.78	74	44.37
23	Torrance	86	71.07	91	101.98
24	Palos Verdes	1309	40.00	259	39.19
25	Carson	376	40.90	123	43.09
26	Long Beach	9	119.97	66	128.13
27	Signal Hill	636	88.15	166	92.19
28	Compton	474	53.85	78	49.89
29	Hawaiian Gardens	70	14.00	95	41.55
30	Cerritos	22	62.49	24	44.39
31	Norwalk	220	89.97	157	65.63
32	Industry	884	31.37	585	18.36
33	Diamond Bar	1366	30.05	973	28.86
34	North El Monte	36	30.33	63	25.12
35	West Covina	31	58.01	37	42.34
36	Glendora	791	29.90	201	17.05
37	La Verne - Azusa	1271	15.81	2391	2.56
38	Altadena	1049	17.24	1954	2.74
39	North Hills - Sylmar	622	40.27	115	21.49
40	Chatsworth	1044	50.53	839	14.80
41	Calabasas	943	23.19	3016	12.14

Model Zone		Developable		Undevelopable	
ID	Name	# of parcels	Value per unit lot size($\$/ft^2$)	# of parcels	Value per unit lot size($\$/ft^2$)
42	Malibu - Point Dume	1095	39.50	2748	10.93
43	Agoura Hills	442	25.46	956	7.69
44	Lake Los Angeles	3213	1.83	33649	1.04
45	Lancaster - Palmdale	14977	6.83	28008	3.35
46	Santa Clarita	12866	5.67	10874	3.43
47	Ventura North County	5316	1.89	3266	1.78
48	Thousand Oaks	4460	10.67	2838	7.58
49	Oxnard - Camarillo	3541	7.72	567	6.93
50	Seal Beach - Los Alamitos	129	71.79	70	98.52
51	Cypress	47	88.08	64	65.74
52	South Buena Park	50	51.92	35	73.05
53	Buena Park - La Habra	414	38.18	484	53.28
54	Placentia	186	86.67	47	84.17
55	Yorba Linda	975	41.82	190	19.15
56	Huntington Beach	624	37.01	314	71.66
57	Garden Grove	113	24.61	114	62.60
58	Anaheim	247	43.92	54	66.49
59	North Tustin	1413	16.88	806	7.64
60	Costa Mesa	73	13.19	111	57.96
61	Santa Ana	275	45.81	265	72.43
62	Tustin	1037	251.05	61	196.48
63	Newport Coast	1721	24.99	1610	31.80
64	Irvine	736	35.36	214	25.83
65	East Orange County	2464	23.09	3860	4.77
66	San Juan Capistrano	2530	52.16	1816	25.46
67	Montclair - Chino	2821	19.61	1095	11.95
68	Ontario	422	47.81	28	44.33
69	Rancho Cucamonga	1488	47.20	74	45.28
70	Upland	3577	32.60	1452	5.45
71	Fontana	2625	33.44	294	29.56
72	Colton	1559	23.21	419	16.84
73	San Bernardino	383	26.84	48	28.25
74	Redlands - Highland	1424	18.62	366	12.19
75	Crestline	2031	13.76	6943	5.76
76	Victorville	16192	4.02	39516	1.80
77	Lucerne Valley	6403	1.97	26351	0.86
78	San Bernardino Mountains	3120	7.93	6924	2.56
79	Northwest Mojave	2744	0.21	15454	0.24
80	Northeast Mojave	7733	0.59	53199	0.32
81	Corona	5208	43.75	168	33.34

Model Zone		Developable		Undevelopable	
ID	Name	# of parcels	Value per unit lot size($\$/ft^2$)	# of parcels	Value per unit lot size($\$/ft^2$)
82	East Riverside	1805	20.49	319	16.77
83	Indio	4954	21.39	1042	23.07
84	East Mojave	3816	1.55	3569	0.74
85	Lake Elsinore	4296	25.06	143	15.28
86	Riverside	114	22.07	86	15.05
87	Moreno Valley	704	28.70	216	19.89
88	Perris	6769	14.49	114	13.40
89	Banning	3117	7.41	688	5.29
90	Hemet	9061	12.21	253	10.83
91	Temecula	4343	19.08	356	18.86
92	Palm Springs	1937	33.41	73	16.19
93	La Quinta	2918	8.59	314	10.55
94	Cathedral City	1125	27.96	49	27.41
95	Palm Desert	1238	25.02	232	27.23
96	Imperial Valley	4971	1.94	407	2.65
97	El Centro	3420	3.58	88	3.66
0	others	5846	9.60	10002	2.67